Comparison of Linear and Nonlinear Approaches on Single Trial ERP Detection in Rapid Serial Visual Presentation Tasks

Yonghong Huang, Deniz Erdogmus, Santosh Mathan, Misha Pavel

Abstract— In this paper, we describe a system for detecting encephalography (EEG) signatures of visual recognition events evoked in a single trial during rapid serial visual presentation (RSVP). In order to investigate the viability of nonlinear approaches in EEG detection and assess the performance comparison, we applied three classifiers (linear logistic regression model, Laplacian classifier, and spectral maximum mutual information projection) in the detection tasks. The EEG was recorded using 32 electrodes during the rapid image presentation (50ms/100ms per image). Subjects were instructed to push a button when they recognize a target image. The results suggest that while the detection of single trial EEG-based recognition is possible, taking advantage of the nonlinear techniques requires data representation that would overcome the non-stationarity of the EEG signals.

I. INTRODUCTION

ELECTROENCEPHALOGRAPHY (EEG) has been a useful non-invasive technique for the assessment and diagnosis of various brain functions and sleep disorders. More recently researchers began investigating techniques in which EEG signals are used to control prosthetic devices and for brain-computer interfaces [1] [2]. Along these lines, EEG has been used to assess the cognitive state of an operator and even to infer whether a human operator detects a target in sequences of images. The latter application requires a system to detect the presence of a small signal during a rapid serial visual presentation (RSVP) of images sequences [3-6].

The problem of searching for targets in vast collections of imagery is one that affects practitioners in a variety of domains – from medical diagnosis to intelligence image analysis. Advances in imaging and storage technology have served to lower the cost of collecting and storing high volumes of imagery. However, the cost of searching through large sets of imagery for important information can often be substantial. In many domains, such as medical diagnosis and intelligence analysis, effective search currently requires the expertise of highly skilled analysts who search through sequences of images in a relatively slow manner. Unfortunately, the availability of skilled analysts is simply insufficient to cope with the volume of imagery to be analyzed. For example, the military reports that most intelligence imagery goes without being examined by analysts [7].

Evoked response potentials (ERPs) arise from coherent neural activity and are reflected in specific morphological changes in EEG waveforms in response to task-relevant stimuli [8]. Prior research demonstrated that ERPs in EEG signals, which reflect the activity of underlying cognitive processes, may be used to identify targets within image sequences presented at very high presentation rates [3-5]. ERPs could be used in conjunction with RSVP of images to dramatically raise the efficiency of searching through high volumes of imagery. During an RSVP presentation, a continuous sequence of images is rapidly presented. A target image in a sequence of nontarget distracter images elicits in the EEG a stereotypical spatiotemporal response.

ERPs are difficult to detect. These signals typically range in amplitude from approximately 1 to 10 μV, while background EEG activity may range from 10 to 100 μV. Common events such as eye blinks or facial muscle activity can completely obscure ERPs. In order to deal with such an inherently low signal to noise ratio, ERP detection has relied on a strategy of trial averaging [9]. Under this strategy, an experimental stimulus is presented to a subject several times. The waveforms elicited by each stimulus are averaged. Background EEG washes out in the averaging process, and the event-induced activity becomes prominent.

While integrating information across repeated presentations of a stimulus is an effective way to identify ERPs, it is an impractical strategy for application domains, such as a triage platform. Repeated presentation of stimuli compromises the efficiency of the search process. In domains where efficient ERP detection is critical, accurate detection of ERPs within a single trial becomes necessary. However, single trial detection of ERPs requires a robust signal processing and classification approach to overcome the problems imposed by the inherently low signal-to-noise ratio.

This work aims to investigate the possibility of using single trial detection of ERPs in the context of a triage platform (identify a subset of images that are likely to contain target images – the triage process trades of specificity for sensitivity). We applied pattern classifiers using these responses to recover spatial components that reflect differences in EEG activity evoked by target vs.
nontarget images. The goal of our technical approaches is to detect ERPs reliably and efficiently.

Various multivariate signal processing algorithms have been proposed for EEG detection [10-14]. Linear techniques are commonly employed in ERP detection. Linear projection [12-14], the current state-of-the-art, served as the baseline approach in this work. We first implemented a linear spatial ERP detector using a logistic regression model to learn an optimal linear discriminator from the spatial distribution of EEG activity (optimal under the exponential parametric model that is assumed). Even though this technique may provide acceptable levels of performance in some situations, it is restricted in their ability to accommodate any nonlinear amplitude and temporal distortions that the ERP waveforms may exhibit from trial to trial even within the same session with the same subject [15]. Such deviations will render the linearity and Gaussianity assumptions invalid, thus will lead to suboptimal detection performance.

Nonlinear techniques have been validated in other signal processing domains, where they have dramatically outperformed linear techniques. Our nonlinear matched filters for ERP detection rely on kernel based projection techniques that have been growing in popularity in the machine learning community [16-18]. The nonlinear techniques were applied in this work to assess the performance comparison among the linear and nonlinear approaches on discriminating EEG activity between target/distracter trials of an RSVP task. We expected our nonlinear approaches would demonstrate identical or greater sensitivity and specificity than state-of-the-art linear ERP detection algorithms without restrictive assumptions about the underlying data.

II. METHODS

A. Data Acquisition and Description

1) Data Acquisition

Subjects were instructed to perform visual target detection amongst distracters. Objects of interest, referred to as targets, consisted of satellite photographs of ships or boats in the midst of a pool of satellite images around a port scene. Both the target and distractor images were drawn from a common high-resolution, broad-area, satellite image. All imagery was presented using the RSVP paradigm (see Figure 1). Images were presented in rapid succession for durations of 50 or 100 milliseconds per image.

EEG data was collected over the course of little over an hour. Each session lasted approximately 20 minutes. There was a rest period of approximately five minutes between sessions. A fixation screen, which lasted several seconds, was used to separate trials. Each trial contained a sequence of approximately 50 images. Of the trials, 50% contained targets while 50% did not. Each trial consisted of a sequence of images in which, if existed, a target image was positioned randomly (except at the first and last 10 images in the sequence).

The data were collected using a 32 channel BioSemi Active Two system. The electrodes were placed on a standard electrode cap, at locations corresponding to the international 10-20 system. A facial electrode was also used to record eye activity. All channels were referenced to a common mean reference. Data was sampled at 256 Hz. Triggers sent by the Presentation script to mark the onset of target and distractor stimuli were received by the BioSemi system over a parallel port and recorded concurrently with EEG signals.

A variety of signal processing components were implemented for reducing the impact of noise artifacts that could compromise ERP detection. EEG was bandpass filtered between 1 Hz and 30 Hz, using an 8th order Butterworth filter to correct for DC drift and limit the effects of 60 Hz electrical line noise. An adaptive linear filter was used to correct EEG signals affected by eye blinks. Using the eye electrode as a noise reference, the adaptive linear filter used the Widrow-Hoff learning rule (least mean squares) to derive an estimate of the impact of eye activity on EEG electrode sites. Once the algorithm had converged, estimates of eye activity at each electrode could be subtracted from the signal at each electrode to decontaminate the EEG signal of eye blink activity.

2) Data Description

EEG data was segmented into epochs. In the case of target trials, each epoch consisted of a two second segment of EEG, one second before, and one second after the onset of target stimuli. For the distracter trials (no target trials), epochs were extracted around the trigger associated with the middle image of each trial block. Data associated with each epoch were stored in a 32*512 matrix (number of channels * EEG samples). The 512 data points represent 256 samples values (i.e., one second of data) before image the trigger and 256 samples after the image trigger. Each epoch served to
provide a picture of spatiotemporal electrical activity across brain regions. Each twenty-minute session yielded
approximately 80 to 90 target epochs and 80 to 90 distracter
epochs each.

We used two sets of pilot data in this work. The first
dataset was collected from one subject. In this experiment,
images were presented at rates of 100ms or 50ms per image
for different blocks of trials. The images either contained
targets or no target. The subject was instructed to indicate
presence of targets by clicking the mouse at the end of each
trial. The subject was required either response as soon as he
had seen a target (button) or respond after a block of trials
had been presented (noButton). There were two sessions.
For each session, there were four sets of data with targets
and four sets of data without targets.

The second dataset was collected from two subjects. It
used the same system and had the same data structures.
There are three sessions for the first subject and two sessions
for the second subject. In each session, there are one set data
with targets and one set of data without targets.

B. Classification Methods

EEG activity resulting from presentation of target and
distracter stimuli was detected by three classification
approaches: Linear logistic regression approach, Laplacian
classifier and Spectral maximum mutual information
projection method.

1) Linear Logistic Regression Classifier

This is the state-of-the-art linear discrimination approach
in ERP detection based on logistic regression [12-14]. The
linear approach relies on the assumption that the EEG
signals are a linear combination of distributed source
activity and zero-mean white Gaussian measurement noise.
Consequently, the optimal ERP detection strategy under this
assumption is to determine optimal linear projections of the
sensor measurements to maximize discrimination ability.

A linear discriminant function is defined as linear
combinations of the components of \( x = [x_1...x_d] \),
\[
y = w^T x + b
\]
where \( w \) is the weight vector and \( b \) is the bias [20]. The
linear projections are optimized using the logistic regression
technique that assumes the conditional class probability
given the projection will follow a logistic model,
\[
p(c|x) = \frac{e^y}{1 + e^y} = \frac{e^{w^T x + b}}{1 + e^{w^T x + b}}
\]
which is consistent with the Gaussian assumption. This
likelihood is parameterized by the weight vector \( w \) and bias
\( b \). The parameters are adjusted by maximizing the likelihood
of the data so that the data matches the logistic model
distribution in (2). In order to compute efficiently, the
iteratively re-weighted least squares algorithm was used to
learn spatial weighting coefficients for discrimination. [21].

2) Laplacian Classifier

This is a classifier operating in a kernel feature space
related to the eigenspectrum of the Laplacian data matrix
[16-17]. The classification rule is based on comparing
angles between test data points and class mean vectors in
the kernel induced feature space. The Laplacian classifier has
demonstrated comparable performance to support vector
machine with less computational complexity, because it does
not require iterative convex optimization [16]. It may
become an alternative to support vector machine for large
data cases.

Consider the two class problem with two class conditional
probability density functions \( p_1(x) \) and \( p_2(x) \). The
classification cost function is derived from the Cauchy-
Schwarz distance between these two probability density
functions, defined as
\[
e_c = -\log \frac{<p_1,p_2>_f}{\sqrt{<p_1,p_1>_f <p_2,p_2>_f}} \geq 0
\]
where \( <p_1,p_2>_f = \int p_1(x)p_2(x)f^{-1}(x)dx \), \( i,j = 1,2 \).
Here \( f(x) \) is the overall probability density function of the
data set given by the mixture of \( p(x) \) with appropriate class
priors. Let \( h(x) = f^{-1/2}(x)p_1(x), g(x) = f^{-1/2}(x)p_2(x) \).
We may denote the argument of the log in (3) as \( L \),
\[
L = \int \frac{h(x)g(x)dx}{\sqrt{h^2(x)dx}g^2(x)dx}
\]
This cost function can be nonparametrically estimated
using a technique similar to Parzen window density
estimation. Given a unimodal Gaussian density and a set of
samples \( \{x_1,...,x_N\} \),
\[
W_{\sigma^2}(x,x_i) = (2\pi\sigma^2)^{-d/2} \exp\{-\|x-x_i\|^2/(2\sigma^2)\}
\]
where \( W \) is the Parzen window or the kernel and \( \sigma \) is
the kernel size. Now we define the matrix \( K_f \) as
\[
K_f(x_i,x_j) = f^{-1/2}(x_i)f^{-1/2}(x_j)K(x_i,x_j)
\]
where \( K(x_i,x_j) = W_{2\sigma^2}(x_i,x_j), i,j=1,...,N \), is the data
affinity matrix (also called the Gram matrix) using a
Gaussian RBF kernel. Note that the increase in the kernel
size is due to the convolution effect of the integrals in (4).
Letting \( K_f(x_i,x_j) = \langle \Phi_f(x_i),\Phi_f(x_j) \rangle \), where \( \Phi_f() \)
is the nonlinear mapping. According to Mercer’s spectral
decomposition theorem, which is also the basic for nonlinear
support vector machines, we have
\[
L = \frac{\langle m_1,f,m_2,f \rangle}{\|m_1,f\||m_2,f\|} = \cos \angle(m_1,f,m_2,f)
\]
which is called the Gram matrix using a
Gaussian RBF kernel. Note that the increase in the kernel
size is due to the convolution effect of the integrals in (4).
Letting \( K_f(x_i,x_j) = \langle \Phi_f(x_i),\Phi_f(x_j) \rangle \), where \( \Phi_f() \)
is the nonlinear mapping. According to Mercer’s spectral
decomposition theorem, which is also the basic for nonlinear
support vector machines, we have
\[
L = \frac{\langle m_1,f,m_2,f \rangle}{\|m_1,f\||m_2,f\|} = \cos \angle(m_1,f,m_2,f)
\]

where \( m_{1f} = (1/N_i)\sum_{j=1}^{N_i} \Phi_f(x_i^j), i=1,2 \), the sample mean
of the i\( ^{th} \) class in the kernel induced feature space obtained
using the samples \( x_i \) that belong to class \( i \).

By utilizing the Parzen window method, the distance
measure between densities in the input space turns into the
distance measure between two classes of data points in a Mercer kernel feature space. In the feature space, the distance measure is the cosine of the angle between the cluster mean vectors. Based on the training data set, we may define the class mean vectors for each class. For the purpose of minimizing the classification cost function, by measuring the angle between a test data point and each of the mean vectors, we can assign the data point to the class that the angle is the smallest.

3) Spectral Maximum Mutual Information Projection

This approach uses the information theoretic concept of mutual information (MI) [22-24] to identify an optimal nonlinear projection using the kernel induced feature space approach [18]. Kernel based transformations provide a way to convert nonlinear solutions into linear ones via a projection into a high dimensional space. The goal is to find a nonlinear subspace projection such that Shannon MI between the projection and the class labels is maximized.

The number of samples in each class is denoted as \( N_c \) and the number of classes is denoted as \( C \). Given a set of data \( \{ x_1, \ldots, x_N \} \) and associated class labels \( \{ c_1, \ldots, c_N \} \). The original data are projected to the kernel feature space through the eigenfunctions according to the theory of reproducing kernels for Hilbert spaces. For the reduced dimensionality \( d \), the projection model can be expressed as

\[
y = V^T \phi(x)
\]

where \( V = [v_1, \ldots, v_d] \) consists of orthogonal vectors, \( \phi(x) \) is the hypothetical embedding vector which consists of the infinitely many eigenfunctions of the kernel \( \{ \phi_1(x), \phi_2(x), \ldots \} \).

The MI between the original feature vectors and the class labels can be expressed as follows [11].

\[
I_S(x; c) = \sum_c p_c E_{x|c} \log \frac{p_{x|c}(x|c)}{p_x(x)}
\]

The probability density functions can be estimated by kernel density estimation [9].

\[
I_S(x; c) \approx \sum_c \sum_{c'} \frac{1}{N_c} \sum_j N_j \log \frac{1}{(1/N)} \sum_j K(x_i^c, x_j^{c'})
\]

Here we use Gaussian kernel functions. The kernel size selection is given by Silverman’s rule [26].

\[
\sigma = \frac{1}{n} \text{tr}(\Sigma_x) \left[ \frac{4}{(2n+1)N} \right]^{1/2}
\]

where \( \Sigma_x \) is the sample covariance matrix, \( n \) is the data dimension, \( N \) is the total number of samples.

According to Nystrom, the kernel feature transformation can be calculated as

\[
\phi(x) \approx N^{1/2} A^{-1} \Phi k(x)
\]

where the eigendecomposition of the kernel matrix defined above \( K = \Phi^T \Lambda \Phi \) yields the necessary parameters for the Nystrom approximation. Then, for the case where one-dimensional projections are sought, equation (10) can be rewritten as

\[
I_S(x; c) \approx \sum_c p_c \sum_{c'} \frac{N_c}{N} \log \frac{\phi^T(x_i^c) \Lambda \mu_c}{\phi^T(x_j^c) \Lambda \mu}
\]

where \( \mu_c = (1/N_c) \sum_{i=1}^{N_c} \phi(x_i^c) \) and \( \mu = (1/N) \sum_{i=1}^{N} \phi(x_i) \). This reformulation of the mutual information maximization problems results in the optimal solution

\[
v = M^1/2 \alpha
\]

where \( M = [\mu_1, \ldots, \mu_C] \), \( p = [p_1, \ldots, p_C] \), \( \mu = Mp \). For the two-class scenario, the optimal \( \alpha \) is given as \( [p_1^{1/2}, p_1^{1/2}] \) as shown in [18]. Combining (8), (12), and (14), we obtain the spectral maximum mutual information projection test statistics for signal detection.

C. Learning and testing procedures

The experimental paradigm for training the classifiers is as follows. Three approaches were conducted for the within-session experiments. The first one is a leave-one-out training and testing procedure [20]. Since leave-one-out artificially induces the assumption of identically distributed training and test data, we employed a different technique to verify generalization. The method is the five-fold crossvalidation procedure. We partitioned each session into five sets of data. One of these five sets was chosen as the testing set and the remaining four as training. The final performance was averaged over the five possible cases for testing data. For the cross-session experiments (which are useful to evaluate the session-to-session transfer of classification performance), we used one full session as the training set and the other session as the test set. The features used for classification are simply the temporal EEG measurements from 32 channels at 512 time instances centered on the target stimuli.

D. Evaluation

Receiver operating characteristic (ROC) analysis [26] is used to quantify the discriminators’ performance. The final performance is assessed using the area under the ROC curve. The actual probability of detection error depends on the frequency of targets, a factor that is determined by specific operational details.

III. RESULTS

In an attempt to investigate the performance comparison using linear and nonlinear detection approaches for ERP discrimination, a complete experimental evaluation was conducted on two different datasets. The goal of these experiments is to determine the effects of nonstationarity on
classifier generalization, as well as to assess the feasibility of rapid image search using the RSVP paradigm.

A. Experiments with One Subject in Two Sessions

The first dataset was collected from one subject with 100ms/50ms image presentation rate and Button/noButton response methods. We analyzed the first dataset through two sets of experiments. In both experiments, we evaluated the Laplacian classifier (LP) and linear logistic regression classifier (LN) with the leave-one-out procedure within the session to investigate the performance comparison between the nonlinear and linear classifiers.

We performed discrimination between targets and distractors for different combinations of presentation rates (100ms or 50ms) and response methods (Button or noButton). The dataset contains one session of each combination (session #1) with 50 samples for 100ms_Button, 49 samples for 100ms_noButton, 36 samples for 50ms_Button, and 46 samples for 50ms_noButton. Also, a second session of each combination (session #2) contains 37 samples for 100ms_Button, 47 samples for 100ms_noButton, 40 samples for 50ms_Button, and 35 samples for 50ms_noButton.

The leave-one-out discrimination performance measured in terms of area-under-ROC for sessions #1 and #2 are shown in Tables 1 and 2. We observe that the Laplacian classifier (LP) has an area of 0.90 to 0.96 (maximum 1.00) while the linear classifier (LN) only has an area of 0.37 to 0.66 in session #1 and 0.91 to 0.97 and 0.46 to 0.67 respectively for session #2. Clearly, the Laplacian classifier produces much better performance than the linear logistic regression classifier for both sessions.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>DISCRIMINATION PERFORMANCE FOR SESSION #1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100ms_Button</td>
</tr>
<tr>
<td>ROC area</td>
<td>0.90 (LP)</td>
</tr>
<tr>
<td></td>
<td>0.37 (LN)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>DISCRIMINATION PERFORMANCE FOR SESSION #2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100ms_Button</td>
</tr>
<tr>
<td>ROC Area</td>
<td>0.91 (LP)</td>
</tr>
<tr>
<td></td>
<td>0.65 (LN)</td>
</tr>
</tbody>
</table>

It was observed that using the leave-one-out technique, the spectral projection classifier performed similarly to the Laplacian classifier, therefore, it is omitted from these tables.

B. Experiments with Two Subjects

The second dataset was collected from two subjects. There are two sessions for the first subject and three sessions for the second subjects. For subject #1, there are 166 samples for session #1 and 174 samples for session #2. For subject #2, there are 168 samples for session #1, 181 samples for session #2 and 159 samples for session #3.

We examined this dataset through two sets of experiments. The aim is to compare the performances among the different discrimination techniques.

1) Five-fold cross-validation within a session

We evaluated the Laplacian classifier with the five-fold crossvalidation technique within a session for two subjects. The average area-under-ROC and its standard deviation are reported. Tables 3 and 4 provide the performances on individual test sets and their statistics for the Laplacian classifier applied to measurements from two subjects. The nonstationarity in the EEG data is evident from the large variation in performance. This also illustrates how leave-one-out testing could lead to artificially inflated performances.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>DISCRIMINATION PERFORMANCE FOR SUBJECT #1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Session#1</td>
</tr>
<tr>
<td>ROC area</td>
<td>0.51 0.81 0.74 0.74 0.82</td>
</tr>
<tr>
<td></td>
<td>mean = 0.72</td>
</tr>
<tr>
<td></td>
<td>std = 0.13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4</th>
<th>DISCRIMINATION PERFORMANCE FOR SUBJECT #2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Session#1</td>
</tr>
<tr>
<td>ROC area</td>
<td>0.89 0.75 0.80</td>
</tr>
<tr>
<td></td>
<td>mean = 0.77</td>
</tr>
<tr>
<td></td>
<td>std = 0.08</td>
</tr>
</tbody>
</table>

2) Generalization across sessions

This study evaluated the session-to session transfer of classification performance. In this experiment, we examine the three classifier performances across sessions. We first investigated training on one session and testing on remaining sessions. The discrimination performance for subject #1 is described in Table 5 and the discrimination performance for subject #2 is described in Table 6. As Table 5 and Table 6 show, for two subjects, a discriminator trained on data from one session generalized well to data from two test sessions which were separated by over a twenty minute gap. The results indicate that three classifiers approach the same performance for across-session performance for both subjects.

<table>
<thead>
<tr>
<th>Table 5</th>
<th>DISCRIMINATION PERFORMANCE FOR SUBJECT #1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>training on session#1</td>
</tr>
<tr>
<td>ROC area</td>
<td>0.87 (LP)</td>
</tr>
<tr>
<td></td>
<td>0.89 (SP)</td>
</tr>
<tr>
<td></td>
<td>0.87 (LN)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6</th>
<th>DISCRIMINATION PERFORMANCE FOR SUBJECT #2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>train on session#1</td>
</tr>
<tr>
<td>ROC area</td>
<td>0.79 (LP)</td>
</tr>
<tr>
<td></td>
<td>0.80 (SP)</td>
</tr>
<tr>
<td></td>
<td>0.79 (LN)</td>
</tr>
</tbody>
</table>
We also examine the performance of three classifiers when trained using data from two sessions and tested on a third session for subject #2. The results shown in Table 7 demonstrate that the three classifiers still perform practically identically. For subject #2, when trained on session #2 and #3, and tested on session #1, all classifiers approached an area under the ROC curve of 0.9. It suggests that we need more data for training to get even better performance.

<table>
<thead>
<tr>
<th>Table 7</th>
<th>Discrimination performance for subject #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROC area</td>
<td>train on session 1#1#2 and test on session#3</td>
</tr>
<tr>
<td></td>
<td>0.86 (LP)</td>
</tr>
<tr>
<td></td>
<td>0.86 (SP)</td>
</tr>
<tr>
<td></td>
<td>0.86 (LN)</td>
</tr>
</tbody>
</table>

IV. CONCLUSION

We studied the effectiveness of three classifiers on single trial ERP detection in the context of RSVP target search in large scale imagery databases. The results confirm that reliable visual target detection in large image databases is feasible with the RSVP paradigm and classification based on EEG measurements. The preliminary results presented here demonstrate area-under-ROC values over 0.80, which corresponds to even higher probability of correct detection for balanced datasets (roughly equal number of target-trials vs nontarget-trials). The main result of this study is that for the limited amounts of data, the two nonlinear classifiers that outperformed the linear classifier on the training sets, did not generalize across sessions.

The results of these experiments demonstrated clearly that the lack of stationarity in EEG [15] is among the issues that need to be addressed properly in feature construction. The leave-one-out technique is generally an inadequate method for biomedical applications especially when small number of samples are utilized (unfortunately this is precisely when researchers tend to use it) due to the artificial performance inflation observed here.

The raw temporal signal-based features (commonly used in ERP detection) coupled with dense EEG arrays yield very high dimensional feature vectors that make it infeasible to expect good generalization given the low sample/parameter ratio. In summary, given the low number of instances in the datasets we have collected, it was preferable to select the simpler linear classifier, however, one also expects the nonlinear classifiers to start becoming desirable in the future when larger training datasets are available for optimizing their parameters.

The future challenges for this field include extraction of fewer more reliable features (perhaps wavelet-based) as well as implementation of real-time continuous detection algorithm that does not require the centering of the input data to the expected position of the target stimuli.

ACKNOWLEDGMENT

This work was supported by DARPA under contract HM1582-05-C-0046 and by NSF under grant ECS-0524835. The EEG data used in the experiments were collected at the Human-Centered Systems Lab, Honeywell, Minneapolis.

The authors would like to thank Robert Jenssen and Umut Ozertem for kindly providing their codes. Special thanks go out to Andre Gustavo Adami for his valuable comments.

REFERENCES


